Is Wikipedia Inefficient? Modelling Effort and Participation in Wikipedia

Kevin Crowston, Nicolas Jullien, Felipe Ortega

Is Wikipedia Inefficient? Modelling Effort and Participation in Wikipedia

Kevin Crowston
School of Information Studies
Syracuse University
Email: crowston@syr.edu

Nicolas Jullien
LUSSI, ICI-M@rsouin
Télécom Bretagne
Email: Nicolas.Jullien@telecom-bretagne.eu

Felipe Ortega
GSyC/Libresoft
University Rey Juan Carlos
Email: jfelipe@libresoft.es

***Abstract*—Concerns have been raised about the decreased ability of Wikipedia to recruit editors and in to harness the effort of contributors to create new articles and improve existing articles. But, as [1], [2] explained, in collective projects, in the initial stage of the project, people are few and efforts costly; in the diffusion phase, the number of participants grows as their efforts are rewarding; and in the mature phase, some inefficiency may appear as the number of contributors is more than the work requires. In this paper, thanks to original data we extract from 36 of the main language projects, we compare the efficiency of Wikipedia projects in different languages and at different states of development to examine this effect.**

I. Introduction

Wikipedia is perhaps the most successful collective knowledge production project ever, having produced more than 3.5 million articles for the English version alone and receiving nearly one million visits per day[1]. The encyclopedia is the product of contributions from thousands of people who give their time and their knowledge to construct articles. As a result, many researchers are closely examining the editing process, seeking insights for managing Wikipedia. These studies have implications as well for organizational projects, as firms are also created to enable collaboration and knowledge integration [3]. The wiki tool seems promising for this application, as it enables distant and sequential collaboration around a structured document [4].

However, despite its tremendous success, there is a growing concern about the global dynamic of the project. In particular, Wikipedia seems to be facing increasing difficulties in recruiting and retaining new editors, so the project is not growing at the same pace than before[2], a situation already noted by researchers [5]. There are many possible explanations for this situation, but the literature on open source software projects [6], and on collective action more generally [1], suggests that such a slow down may simply be the result of the project entering a mature phase in which it needs fewer additions and thus fewer contributors.

However, a more troubling possibility is that the evolution of Wikipedia has led to the development of processes that make contributing to these projects more difficult, with the result that the efforts of editors is increasingly spent on work other than articles. Such inefficiencies would have both a direct impact on the output of the project as well as an indirect impact by making the work less rewarding [7] and so reducing the number of active editors.

To distinguish these possibilities, we have to better understand how production is organized in Wikipedia, and to distinguish a slowdown due to a reduction in contributors from a decrease in the efficiency of article production, in other words, to look closely at the project's production function, in terms of the number of articles produced compared to the number of people available to contribute. Specifically, we propose to compare the main Wikipedia language projects regarding the efficiency of this conversion of inputs to outputs. Evaluation of the efficiency of production of articles does not seem to have been done for Wikipedia.

The analysis is facilitated by the structure of Wikipedia. Wikipedia is not a single project, but rather a multitude of independent sub-projects. For each language there is a separate version of the encyclopaedia with its own editor community and collection of articles. Importantly, the projects are at different levels of maturity, some quite mature, others still getting started and others somewhere in between. However, the projects all share the same tool for collaborative edition (MediaWiki) and the same basic rules for collaboration, the "five pillars" of Wikipedia[3]. As well, the global structure of the projects, measured as a network, the nodes being the articles and the links the links between the articles, seems to be about the same in terms of "degree distributions, growth, topology, reciprocity, clustering, assortativity, path lengths, and triad significance profiles", at least for the main projects [8]. In contrast to studies on open source software (see for instance [9], [10]) that compare project that use various technologies, programming languages and collaborative tools, this uniformity may help us to better understand, in their difference, what differences are due to process evolution.

The paper is organized as follows: first, we define the inputs and the outputs which are to be evaluated. We then describe our analysis approach, multiple-input multiple-output efficiency techniques (specifically data envelopment analysis), and present our hypothesis on the performance of the projects.

[1]For statistic on Wikipedia, visit http://stats.wikimedia.org/EN/.

[2]See for example http://en.wikipedia.org/wiki/Wikipedia:Areas_for_Reform and http://meta.wikimedia.org/wiki/Research:Index.

[3]http://en.wikipedia.org/wiki/Wikipedia:Five_pillars.

Finally we present the data and our current results and compare to prior work. We discuss these results in a conclusion section.

II. Theory development

Defining itself as an *"online encyclopedia"*, incorporating "*elements of general and specialized encyclopedias, almanacs and gazetteers*", Wikipedia covers a large scope. For each entry or article, it aims at *"explain[ing] the major points of view in a balanced impartial manner"*, with *"verifiable accuracy"* and *"references"*. It is produced by volunteers who interact, knowing that *"anyone can edit, use, modify, and distribute"*[4].

The voluntary nature of contributions does not mean that there is no organization or rules regulating the way people may contribute. On the contrary, as [12] show in an analysis of the English project, the rules governing Wikipedia are increasing in number and in complexity. The process of creation is regulated by numerous social conventions and unwritten rules [13]. Furthermore, there are different type of contributors; a few are "power editors" [14]; a few have official responsibilities in the project and special powers[5]. Thus, in many aspects, Wikipedia seems to follow the knowledge commons framework proposed by [15].

We are interested in particular in the efficiency with which the different language Wikipedias turn inputs into outputs. To do so, we must first identify which specific factors to examine as inputs and outputs. We consider inputs and outputs in two distinct processes.

First, to be successful, Wikipedia must recruit editors. Research has shown that a mix of experienced editors and fresh newcomers increase the likelihood for an article to reach a good level of quality (a "Feature Article") [7], [16]. For the recruitment process, we take as input the number of potential contributors, namely those who speak the appropriate language, have access to the Internet and have a suitable level of education to be good contributors. We consider the appropriate educational level to be the tertiary level, as [17, table 6] and a survey on the French Wikipédia [18] showed that the Wikipedia contributors are significantly more educated than the readers. The output of the recruitment process is the number of editors (of different types, described below) contributing to the project.

The second process we examine is the creation of articles. As far as the input are concerned, the main input to Wikipedia is the effort of the contributors, that is, the output of the first process. We used the number of edits performed as an initial measure of output, as edits are the most immediate output of the work of the editors. But the real output of Wikipedia is the articles produced, so we consider as another output the number of new articles created. Of course, articles differ in length and quality. Wikipedia has several defined quality levels for articles, from "article needing to be improved" to "featured articles" (FA or a comparable phrase in the various languages)'[6]. FA is one of the few measures of quality articles shared by every language project, so to integrate the notion of quality, we consider the number of new featured articles created in a period of time. We consider that the number of FAs in a project reflects the projects's interest in increasing the quality of the articles, instead of, or in addition to just the quantity (the number of articles).

Having identified the inputs and outputs of the two processes, we turn to the link between them. Economists formalize the link between inputs and output as a production function [19]. To be efficient is to reach the maximum possible outputs for a given amount of inputs. In our case, the form of this function is unknown, as are the coefficients relating its components. However, we are not trying to propose a characterization of the Wikipedia production function, but to evaluate if some projects are more (or less) efficient than the others. Since the seminal work of [20], this assessment can be done by looking at the "frontier production function", which describes, for various combination of inputs and outputs, the producers who are efficient, i.e., the ones for which none of the outputs can be increased, without either or several of the inputs increasing or other outputs being reduced, and vice versa.

An additional consideration in analyzing the efficiency of production is the question of "return to scale", that is, whether a big project may be more efficient because of its size. Return to scale may be positive because the knowledge production in Wikipedia is in a way cumulative: the production of new articles or the improvement of existing ones depends on the existing stock of articles, both in terms of quantity and quality. This aspect of production is consistent with other knowledge production functions. It is assumed in the literature that the new knowledge is positively correlated with the existing stock of knowledge, even if the form of the function is still in debate (see [21] for a review of the literature on the knowledge production function).

In the case of software, however, it has been found that productivity decreases after a certain growth. This decrease is explained by the fact that software production can be considered as a problem solving activity, where the number of problems is finite. So, as time goes, the number of problems remaining is decreasing, when the number of contributors is not always, leading to a negative return to scale. However, it is not clear if this result will apply to Wikipedia: the fact that the people may have a growing knowledge of how to contribute may lead to a positive impact, as well as the fact that there are few articles in common between the various projects, which indicates there is still lots of articles to write in each language compared with the English version[7].

Drawing on the framework developed above, in this paper, we examine two specific research questions regarding partic-

[4] All the citations of this paragraph come from the Wikipedia page describing its fundamental principals, or "five pillars": http://en.wikipedia.org/wiki/Wikipedia:Five_pillars. See also [11] for a discussion of Wikipedia as a model for collaboration.

[5] http://en.wikipedia.org/wiki/Wikipedia:About

[6] http://en.wikipedia.org/wiki/Wikipedia:Article_development

[7] [22] studied 25 language projects and found that articles present in all projects represent only 1% of the total, when 74% of the articles were present in one language only.

ipation in Wikipedia projects and on the production of edits and articles in the main Wikipedia language projects. First, we examine the comparative efficiency with which different language Wikipedias recruit editors from the available pool of potential editors. We will first make the hypothesis that there is constant return to scale in this recruitment, as the growing popularity of the project may compensate a decreasing interest for participating in the population. We then relax this hypothesis and discuss the results.

The second hypothesis concerns the form of the production function of edits and articles. The form of the production function in open source software, and the observation that the individuals most involved in Wikipedia actually contribute less in proportion [23]–[25] seems to indicate a decreasing return to scale for the older and bigger language projects. This result may lead to a measured lack of efficiency in the bigger projects taking into account only the production of new articles. However, if these projects trade a decrease in the production of new articles for a more quality-oriented production, this effect should be visible in FA production.

III. Data and analysis

The unit of analysis for our study is the various language Wikipedias (e.g., French, German, Japanese). As the English Wikipedia is an outlier in many ways (oldest, largest, etc.) we decided to not include it in this analysis as we were concerned that it would have too great an influence on the results.

A. Data

We describe in turn the data we obtained to measure the inputs and outputs to the recruitment and edit and article production processes.

1) External data: Reliable data is hard to find for studies performed at the national or super-national level. We therefore had to estimate several of our measures by combining published statistical data obtained from various external sources.

To estimate the input to the recruitment process, we needed data on the number of potential editors for each language Wikipedia, which we took to be the number of people who speak the language, have access to the Internet and who have a tertiary education. To estimate the Internet population, we took the data from Internet World Stat[8]. This site aggregates Internet usage data from several sources, including "data published by Nielsen Online, by the International Telecommunications Union, by GfK, local Regulators and other reliable sources"[9]. Some data are available at language level[10], but the Internet population had to be calculated for others[11]. For these, we calculate the total number of users taking the internet rate in the main country(ies) of speaking times the population of this(ese) country(s) and the population of the minority speaking the language times the internet rate in the others countries where the language is spoken.

We do the same regarding the number of people with a tertiary-level education by language. The primary data for this measure comes from UNESCO[12]. Of course, these sources provide only an approximation of desired input variables, but are our best estimates. Still, drastic inaccuracy in the estimates would affect our estimates of the productivity.

2) Wikipedia data collection: As did prior studies of Wikipedia [23]–[28] we relied on Wikipedia internal data to estimate the number of people involved, their characteristics and level of activity. To compute these variables, we obtained the complete database dump with all edits performed in 39 Wikipedias in different languages. Table I shows the list of all languages that have been included in this study. These dump files include all required metadata to trace the creation of new articles and individual changes on any page in these Wikipedia projects. In Wikipedia terminology these edits are known as revisions. We were able to retrieve from the dump files the metadata describing each revision, including the identifier of the user who made that edit, its timestamp or the identifier and title of the page that was edited. Data were obtained for each language project for the month of August 2011.

For each language we also retrieved an additional file containing information about any special privileges granted to certain Wikipedia users. For instance, in this way we can identify administrators, as well as bots (software programs using Wikipedia accounts to perform routinary or targeted changes in an automated way).

By examining which user performed each revision, we obtained counts of the number of editors actually active in the month. We separated these by the number of edits performed in to three groups, following the definitions used by the Wikimedia Foundation[13]: very active Wikipedians (Wikipedians who contributed 100 times or more in this month); active Wikipedians (Wikipedians who contributed 5 times or more in this month) and other contributors. We wanted to count these groups separately because they make such different levels of contribution that grouping them together could give misleading results. As well, the seminal work by [29] shows that these different levels of participants are needed to build the encyclopedia.

The main focus in this article is the study of effort spent by human editors that can be univocally identified in any of these Wikipedia languages. As a consequence, we first filtered out all revisions performed by anonymous editors, since only their IP address is recorded in that case, which is not valid to identify individuals, as many editors can share the same IP address.

We also elided all revisions undertaken by bots, as they

[8]http://www.internetworldstats.com/

[9]http://www.internetworldstats.com/stats.htm

[10]Chinese, Spanish, Japanese, Portuguese, German, Arabic, French, Russia, Korea.

[11]Dutch, Hungarian, Persian, Romanian, Bulgarian, Croatian, Greek.

[12]UNESCO: Educational attainment of the population aged 25 years and older / Latest year available, http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx, except for Russia and China, which come from the OECD (http://www.oecd.org/dataoecd/22/10/40111027.pdf and http://browse.oecdbookshop.org/oecd/pdfs/free/9109031e.pdf respectively).

[13]http://stats.wikimedia.org/EN//

do not represent work done directly by editors, which is the production process we wanted to study.

We also used the revisions to identify the number of new articles and new FAs created in the month. For every revision the whole text reflecting the state of the page after the application of changes is also available. We took advantage of these data to search for the templates used in each of these 39 languages to award the Featured Article (FA) status to Wikipedia articles of exceptional quality, using regular expression patterns on the parsed texts. As a result, we could track the FA status of any Wikipedia article along their whole history, including periods in which the article may have lost this special status for some time.

The data extraction has been implemented as a software program written in Python to automate this process. This program is part of WikiDAT (Wikipedia Data Analysis Toolkit)[14]. This is a multi-purpose framework that combines Python, R and MySQL with the aim of facilitating Wikipedia data analysis for any of the 280 languages currently available in the free encyclopedia. The use of Python lxml[15], an efficient library for XML parsing, and multiple sub-processes, let us speed up significantly data retrieval, extracting and computing all metadata and additional information described above (for instance, as far as the English Wikipedia is concerned, 444,946,704 revisions in 27,023,430 pages were analyzed in approximately 44 hours). This massive data analysis allowed us to develop more precise data than those presented by the Wikimedia Foundation[16] as far as the edits and the contributors are concerned, and to include new, original data, the number of FAs and the number of new FAs by month.

B. Analysis approach: DEA modelling

There are several techniques for estimating the frontier production function. A detailed comparison is out of the scope of this paper, but see [30] for a discussion of these techniques regarding software production. We choose to use Data envelopment Analysis models originally proposed by [31]. Data Envelopment Analysis is a 'data-oriented' approach for evaluating the performances of a set of peer entities, called Decision Making Units, or DMUs in the original source, but in this article, each Wikipedia language project. According to the definition of relative efficiency, a DMU "is to be rated as fully (100%) efficient on the basis of available evidence if and only if the performances of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs" [32, def. 1.2, chapter 1, p. 3].

We decided to use DEA following [10]'s use in the case of open source software: "these models were developed to measure the efficiency of non-profit units, for whose inputs and outputs no clear market prices exist and also no clear evaluation relations" (p. 403), and "DEA is a non-parametric optimization method for efficiency comparisons without any need to define any relations between different factors or a production function. In addition, DEA can account for economies or diseconomies of scale, and is able to deal with multi-input, multi-output systems in which the factors have different scales" (p. 398).

Two main criteria have to be taken into account regarding the choice of a DEA model: the orientation of the model (input-oriented or output-oriented) and the return to scale in the production process.

Regarding the first criteria, as in [10], an output-orientation seems to be more appropriate, as, for a period of time, the inputs (the volunteers in an open online project, the Internet population) are more or less fixed and the goal is to maximize the output.

Considering the second criteria, based on the study of [1] on collective action, on the analysis of software projects, and on the discussion above, it seems rather difficult to assume a constant return to scale. Instead, these projects seem to have a increasing return to scale in a first phase, and then a decreasing one. So we will choose a model allowing to add a variable to control the return to scale, the BCC model, and more specifically the BCC-O (output oriented) model [33].

Theorem 1 by [34] gives the following condition for the sense of the return to scale:

1) Increasing RTS prevail at $(\hat{x}_0,\hat{y}_0)$ if and only if $v_0^* > 0$ for all optimal solutions.
2) Decreasing RTS prevail at $(\hat{x}_0,\hat{y}_0)$ if and only if $v_0^* < 0$ for all optimal solutions.
3) Constant RTS prevail at $(\hat{x}_0,\hat{y}_0)$ if and only if $v_0^* = 0$ for at least one optimal solution.

For the data analysis, we used [35]'s macro under SAS, with non-constant return on scale constraint. The original program is an input-oriented one, so we had to change the equations into an output-oriented one (equations 4.27 to 4.30 in [36, p. 89]).

C. Model

To test the first research question, we took as inputs the Internet population and count of people with a tertiary education speaking each language, and, as outputs, the count of the various types of contributors (very active Wikipedians, active Wikipedians and other contributors) for the different language projects. We then examine the relative efficiency of the projects in transforming the population of potential contributors into contributors at different levels of activity.

To test the second research question, we took as inputs the outputs of the first model, i.e., the count of the various types of contributors (very active Wikipedians, active Wikipedians and other contributors). To control for the size of the project and the knowledge already available, we also take as inputs the number of existing articles and the number of existing links. We considered also adding the stock of FAs as an input. However, since the number of FA is very low compared to the total number of articles, we thought that it would be unlikely to have much influence on efficiency. We checked this point re-runing the models presented below adding the stock of FAs

[14]http://libresoft.es/node/564
[15]http://lxml.de/
[16]http://stats.wikimedia.org/EN//

Fig. 1. Number of contributors versus Internet population.

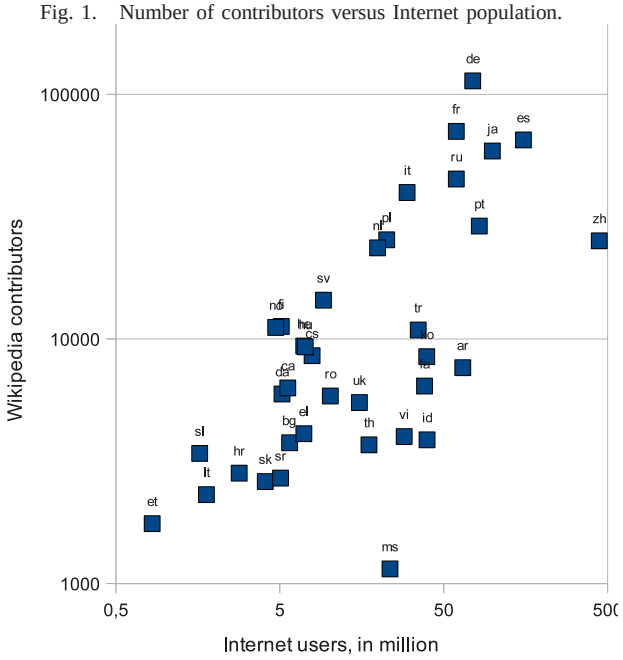
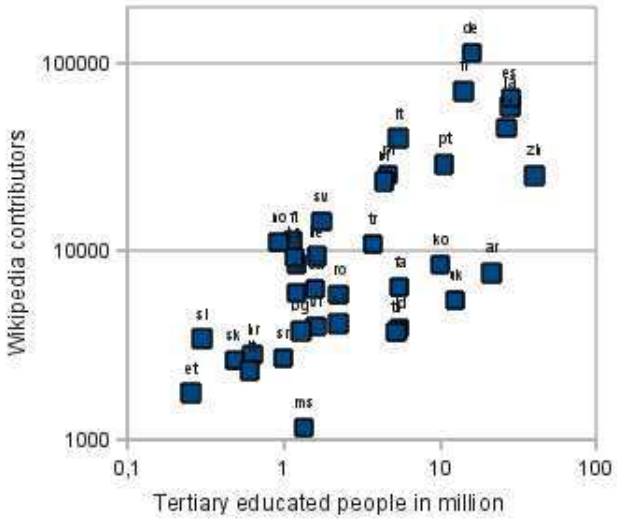


Fig. 2. Number of contributors versus population with a tertiary education.



as an input. As expected, we obtained nearly the same results. For simplicity therefore, in the rest of presentation we will only present models without the FA stock as an input.

As an output, we used first the number of edits per month as a measure of the level of activity of the project. In a second model, we examined how the process transforms editors' contributions into articles, while controlling for possible differences in quality. We used as an output the number of new articles produced per day during the month, and then the number of new articles along with the number of new FAs.

IV. Findings

In this section, we present our findings about the inputs, outputs and efficiencies of the two processes.

A. Recruitment

Plotting the data shows a strong (but not perfect) correlation between the total number of Wikipedia contributors and the Internet population (figure 1), and the total tertiary-educated people (figure 2). Using the DEA model, we are able to determine the different levels of efficiency in the conversion of these inputs to the Wikipedia community of contributors. As said before, we first applied a constant return to scale model, the we introduced the possibility of a variation.

The results for this analysis are shown in Figure 3. The projects are listed in decreasing order of size). The bars indicate relative efficiency. The longest bars, representing 100% efficiency, are for projects that are on the efficient frontier, creating the most outputs from the particular mix of inputs. Shorter bars represent projects that use a similar mix of inputs but produce less outputs than other projects. The results indicate varying levels of efficiency in converting potential editors to actual editors. Specifically, language projects such as Malaysian (ms), Arabic (ar) and Chinese (zh) have many fewer editors than would be suggested by the population of Internet users who could become editors, while Estonian (et), Hungarian (hu), Norsk (no) and Finnish (fi) show high efficiency in recruiting editors.

As far as the return to scale is concerned, table I presents the sign of the return to scale variable, v_0. It seems that the biggest efficient projects have entered in a decreasing return to scale phase ($v_0 < 0$), suggesting increasing difficulty to recruit new Wikipedians. In the other hand, the smaller projects, when they are efficient, seem to be still in an increasing return to scale phase.

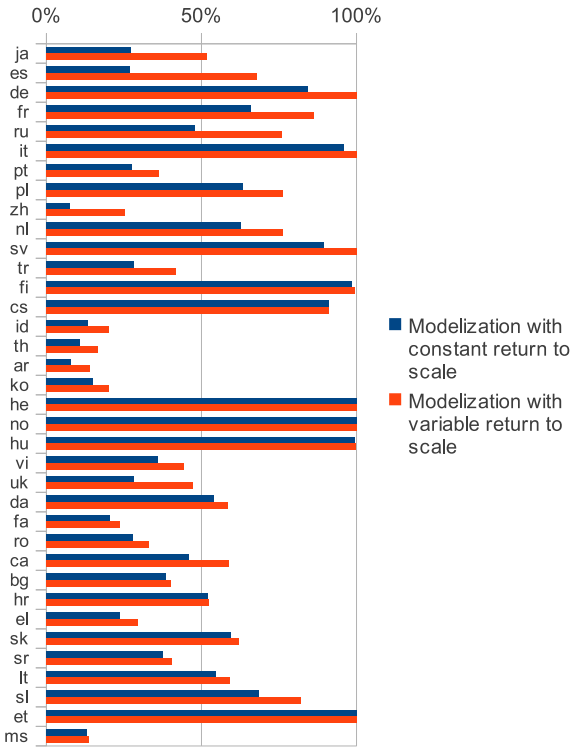
B. Production

The second model examines the production of edits to articles, of new articles and of new articles and new FAs. The results are shown in figure 4 (production of edits, articles and FAs). Yellow bars show the efficiency of producing edits, navy blue bars, efficiency in producing new articles, and red bars, efficiency in producing articles and new FAs.

The difficulties of the main projects to maintain a constant level of activity as the stock of articles increases appears clearly, as the return to scale is systematically negative for the larger projects (see table II). But beside the Japanese project, the main projects are still efficient in terms of level of activity. On the other hand, projects that apparently find it difficult to recruit editors may still be efficient in the converting the effort of the workforce available into edits and articles, as is the case for the Malaysian (ms) and Farsi (fa) language Wikipedias.

Figure 4 helps to explain if edit activity is due to a high level of production of articles or the results of activities that consume activity but do not lead to an increase of the stock of article, either positive, such as a focus on improving existing

Fig. 3. Efficiency in recruitment of contributors.



Note: projects are listed in decreasing order of size.

TABLE I
RETURN TO SCALE FOR THE RECRUITMENT OF CONTRIBUTORS.

Japanese	ja	-1.57
Spanish	es	-1.60
German	de	-0.04
French	fr	-0.11
Rusian	ru	-0.12
Italian	it	-0.12
Portuguese	pt	-0.17
Polish	pl	-0.14
Chinese	zh	-0.15
Dutch	nl	-0.10
Swedish	sv	-0.45
Turkish	tr	-0.29
Finnish	fi	-0.03
Czech	cs	-2.19
Indonesian	id	-0.65
Thai	th	-0.38
Arabic	ar	-0.73
Korean	ko	-0.09
Hebrew	he	-0.08
Norwegian	no	0.02
Hungarian	hu	-0.14
Vietnamese	vi	-0.36
Ukrainian	uk	-0.64
Danish	da	-1.31
Farsi	fa	-0.62
Romanian	ro	-0.15
Catalan	ca	-0.78
Bulgarian	bg	-0.49
Croatian	hr	0.34
Greek	el	-0.75
Slovak	sk	0.42
Serbian	sr	-0.16
Lithuanian	lt	0.13
Slovenian	sl	0.15
Estonian	et	-0.19
Malaysian	ms	-0.56

In red when the project is efficient.

articles, or negative, such as bureaucratic discussion or even edit wars. For instance, the high level of edits in the French (fr) and the German (de) language projects seems to be due to a focus on FA production rather than the production of new articles, for which those projects seem rather inefficient (even after having taken into account the decreasing return to scale in the model). On the other hand, projects of intermediate size, such as the Rusian (ru) or the Italian (it) Wikipedia, are still very active on the level of new article production.

Finally, the level of edits seems to be a good indicator of the level of final production, as few of the projects which are inefficient at the edit level are efficient at the article or article and FA level. The exception to this rule are the Lithuanian (lt), Portuguese (pt), Polish (pl) and Indonesian (id) projects. A possible explaination, given by [37], and proven for the Indonesian project by [38], is that an important part of the articles in those projects are directly translated from English, and these articles require less editing to be published.

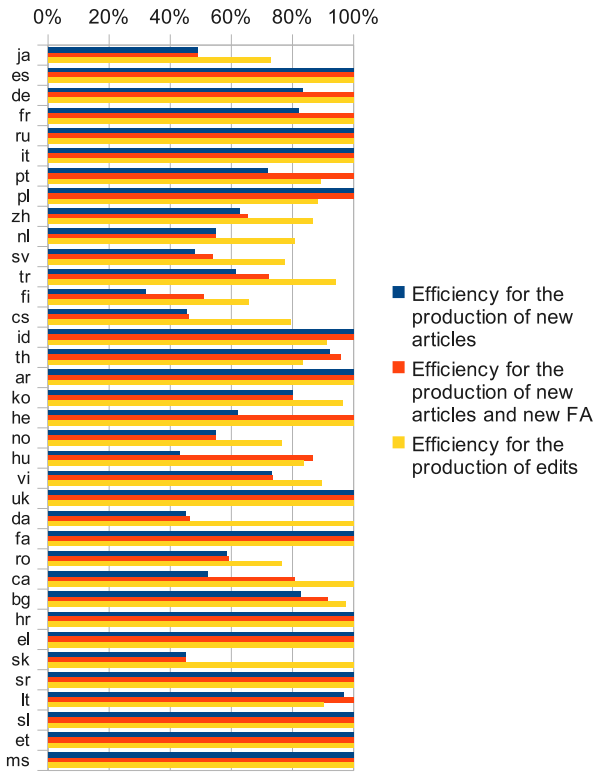
V. DISCUSSION

Our analysis shows striking differences in efficiency in the two processes among the projects. For the differences in efficiency in recruiting participants, the size of the project seems to matter, as all the larger projects are assessed as being inefficient. In the model adding a factor for return to scale, the larger projects increase their performance, with a negative return to scale (V_o being negative). In other words, it may simply be that the largest projects have reached a size where it is harder to make a new contribution and so harder to recruit new Wikipedians.

Nevertheless, there remain striking differences in efficiency among the smaller projects. We propose two possible explanations for these differences. First, many of the less efficient projects have a lower level of tertiary-educated people compared to the efficient group. This difference could be a key to explaining the low efficiency of recruitment. A second hypothesis is on the control of the information: many of the low-efficiency projects are tied to countries where the Internet and the production of information is more closely controlled by the authorities than in the efficient group. It may be that freedom of expression is pre-requisite for efficient recruitment of editors. [39]'s recent study on the Chinese Wikipedia gives arguments for this hypothesis.

As for production, it appears that some of the difference can be attributed to the level of maturity of the projects. Newer projects have fewer articles and so it is easier for contributors to find topics that have not been covered. For the larger and older projects, is the gap between efficiency in editing and in creating new articles because work is being

Fig. 4. Efficiency for the production of edits and articles (new articles and new FAs).



Note: projects are listed in decreasing order of size.

TABLE II
RETURN TO SCALE FOR THE ACTIVITIES.

Projects	Number of edits	Production of new articles	Production of new articles and of new FA
ja	-0.02	-0.95	-0.95
es	-0.01	-0.36	-0.36
de	-0.99	-1.20	-0.93
fr	-0.49	-1.22	-0.98
ru	-0.05	-1.00	-0.12
it	-0.17	-0.57	-0.47
pt	-0.10	-0.80	-0.89
pl	-0.10	-0.67	-0.71
zh	-0.03	-0.40	-0.33
nl	-0.13	-1.01	-1.01
sv	-0.22	-1.43	-1.26
tr	-0.25	0.13	-0.82
fi	-0.10	0.34	0.10
cs	0.10	0.69	0.25
id	-0.05	0.21	0.42
th	0.81	1.43	1.32
ar	-0.36	0.00	-0.59
ko	0.02	0.32	0.32
he	-0.09	0.58	0.12
no	-0.21	0.15	0.15
hu	-0.03	0.04	0.10
vi	-0.06	0.19	0.20
uk	-0.01	0.15	-0.34
da	-0.10	1.76	1.63
fa	0.01	0.21	0.06
ro	-0.22	0.28	0.35
ca	-0.18	0.22	-0.26
bg	0.20	2.76	1.22
hr	2.77	3.23	2.77
el	0.26	2.12	1.34
sk	0.19	2.61	2.61
sr	-0.56	0.53	0.45
lt	0.49	0.80	0.22
sl	-0.10	3.75	0.94
et	0.77	1.56	1.90
ms	0.65	1.40	2.03

In red when the project is efficient.

directed to improving the quality of the article, or is it a sign of inefficiency (ineffective edits)? The current evidence is inconclusive. [26], [28], confirmed by [5], found that after taking into account age and visibility (using Pagerank as a proxy), FA status could be predicted by an increased number of edits or number of editors. [40] found that in 2001, 90% of edits were done in the Main namespace on the English Wikipedia but that this number dropped to 70% by June 2006, suggesting that the efforts of editors are being diverted to less productive activities. However, [28] found that articles with more discussion on their Talk page were generally ranked higher in quality according article ratings, suggesting the tradeoff between simple production and efforts to improve quality.

A. Limitations

The major limitation of our study is that the validity of our analysis is dependent on the quality of the data used. We are quite confident in the data extracted from the Wikipedia dumps. However, a limitation of the work presented here is that we evaluated the projects on a single month, August 2011. Having only one month of data could lead to misinterpretations, especially taking into account that it can be vacation month in some countries. We are working on extending the analysis on twelve months and doing a mean estimation of the efficiency of the various projects.

On the other hand, the external data used for the inputs to the recruitment process are only best estimates. Systematic errors in these data would affect our measure of the relative efficiency of recruitment for the affected languages.

The most significant limitation to our estimate of the efficiency of the edit and article production process is that we have only a partial information on the input: the number of people involved, but not their effort, as we do not know, for instance, how many hours each person spend on the project. As we want to compare Wikipedia language projects, we can only assume that from one project to another, the mean time spent is the same for each type of contributor. Violations of this assumption will affect our measure of the relative efficiency of the projects.

Another limitation in terms of production measurement is the choice of not incorporating the anonymous contributions. The share of such contributions varies dramaticaly between the different languages (between 26% for the Japanese project and

TABLE III
SHARE OF THE EDITS DONE ANONYMOUSLY, IN TOTAL AND IN AUGUST, 2011.

Projects	% of anonymous edits	
	in the project	in August 2011
ja	36.1	26.5
es	26.4	24.1
de	18.6	11.6
fr	14.0	11.4
ru	16.7	17.7
it	17.9	19.0
pt	19.9	22.1
pl	13.3	8.3
zh	12.8	16.4
nl	9.8	8.7
sv	13.8	11.2
tr	23.9	17.8
fi	15.4	12.9
cs	9.3	8.8
id	8.9	8.6
th	14.8	15.0
ar	12.0	9.2
ko	14.6	15.9
he	9.7	6.9
no	8.2	5.7
hu	5.4	3.8
vi	8.3	7.4
uk	5.2	4.2
da	10.2	6.1
fa	5.1	2.9
ro	6.7	4.9
ca	5.4	2.8
bg	13.7	13.1
hr	8.3	6.5
el	13.7	12.4
sk	6.4	4.3
sr	5.1	4.3
lt	6.9	4.3
sl	3.9	1.7
et	7.6	8.1
ms	7.4	3.7

less than 2% for the Slovenien one, see Table III for details), a variation which could explain, in part, the differences in efficiency between the projects. For instance, the Japanese case suggests that having a large proportion of the edits made by anonymous contributions does not result in efficient additions to the articles.

With regard to outputs, a final limitation is that we had only the count of FA projects to assess the quality dimension of the articles. We note that there are concerns about the validity of FA status as a quality measure. The concrete rules of the process of granting FA status to a Wikipedia article varies between the different language projects, but in general it involves voting on the quality of the article: the article should receive a substantial proportion of positive votes to be granted the FA label. However, [41] showed that the argument of quality to qualify an article as FA varies from one language to another. Furthermore, when external expertise is mobilized to evaluate the quality of FAs, as in [42], they show strong variations regarding their assessed level of quality.

B. Future research

The work presented here lays the groundwork for additional research. First, future work should include outputs along additional dimensions, considering factors such as article size and quality, as well as the whole organization of the encyclopedia, which are the usual dimensions for analyzing documents in library studies, see for instance [43] on a comparison of Wikipedia with other encyclopedias.

For this purpose, we will require a better measure of article quality. The most comprehensive attempt to develop criteria to judge article quality may be the ones by [44] and [45]. [44] looked at the information quality process both in the organization (number of editors, of edits, ratio between edits in talk pages and in content pages, etc.), and in people's interaction (via a content analysis of a set of feature articles' talk pages). [45] extend these criteria to 13 criteria (see p. 126 for the complete list), drawn from data analysis (length of the article, existence of references, etc.) but also human (expert) evaluation of the quality. They show a correlation between these criteria and the rank in search engine, with a good correlation but a strong dispersion. In both cases, the automation of the methodology to a whole project, not to say to different languages, seems impossible. There are efforts to automatically analyze the articles, but these are currently not yet enough effective to be of use [46]. Indeed, even the fact that an article is a FA is not coded in the projects' data base, and instead has to be extracted revision by revision from the text of each article in the projects' dump.

Second, this analysis would benefit from distinguishing more finely among different kinds of editors, beyond the three levels of very active, active and other used here. It has been shown that the number of article per authors follows a power law [27], as does the number of contributions per author [5], [23], meaning that there are a small number of editors who make disproportionate contributions. [23]–[25] showed that the percentage of contribution coming from power editors is decreasing over time, which may explain the reduced efficiency of the larger projects.

On the flip side, the analysis should also consider the contributions of non-registered editors. [47], examined contributions from registered and non-registered users, showing the importance of anonymous contributors in the total production.

Finally, moving down from the level of an entire language project, this kind of analysis might also be done at portal or subject level. [48] does a similar analysis though only on a small subset of article. They obtain counter-intuitive results, as it seems from their analysis that the subject having the more feature articles (high-density subjects in his terminology) have longer articles, but fewer edits and contributors than the low-density subjects, while the ratio between major and minor edits is the same in the two groups. It seems also that there is more often a single major editor in the high-density subject articles. This result have been confirmed by [49] who found a positive impact on an article's quality from an increase in the size of the number of editors only when a small core of editors

performed the majority of editorial work.

VI. Conclusion

The results of our analysis are suggestive, but clearly just a first step. The work presented here provides an initial step to identifying difference in the work practices of the various Wikipedias, shedding light on an important example of cultural variety on the practicing of collective intelligence, and proposes a way to extend the work initiated by [41], [50], and [51]. However, while we have shown differences in efficiency, we do not yet fully understand why these differences arise. The next step of the research will be to find better explanations for these differences and to test the possible explanations offered above. Better understanding these differences should provide insight for managing both Wikipedia as well as other open knowledge creation projects.

References

[1] G. Marwell and P. Oliver, *The Critical Mass in Collective Action: A Micro-Social Theory*. Cambridge University Press, Cambridge, 1993.

[2] P. Oliver, G. Marwell, and R. Teixeira, "A theory of critical mass interdependence, group heterogeneity, and the production of collective action," *American Journal of Sociology*, vol. 91, no. 3, pp. 522–556, November 1985.

[3] R. Grant, "Toward a knowledge-based theory of the firm," *Strategic Management Journal*, vol. 17, no. Winter, pp. 109–122, 1996.

[4] H. Hasan and C. Pfaff, "The wiki: an environment to revolutionise employees' interaction with corporate knowledge," in *Proceedings of the Australasian Computer-Human Interaction Conference*. Sydney: OZCHI 2006, 2006.

[5] F. Ortega, "Wikipedia: A quantitative analysis," Doctoral Thesis, Universidad Rey Juan Carlos, Móstoles, Spain, Apr. 2009.

[6] S. Koch, "Effort modeling and programmer participation in open source software projects," *Information Economics and Policy*, vol. 20, no. 4, pp. 345–355, 2008.

[7] S. Ransbotham and G. C. Kane, "Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in wikipedia," *MIS Quarterly*, vol. 14, no. 9, pp. 541–545, September 2011.

[8] V. Zlatić and H. Stefancić, "Model of wikipedia growth based on information exchange via reciprocal arcs," *EPL*, vol. 93, no. 5, p. 58005, 2011.

[9] K. Crowston, J. Howison, and H. Annabi, "Success in free and open source software development: Theory and measures," *Software Process Improvement and Practice*, vol. 11, pp. 123–148, 2006.

[10] S. Koch, "Exploring the effects of sourceforge.net coordination and communication tools on the efficiency of open source projects using data envelopment analysis," *Empirical Software Engineering*, vol. 14, pp. 397–417, 2009.

[11] J. M. Reagle Jr., *Good Faith Collaboration. The Culture of Wikipedia*. Cambridge, Massachusetts: MIT Press, 2010.

[12] B. Butler, E. Joyce, and J. Pike, "Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in Wikipedia," in *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 1101–1110.

[13] F. B. Viégas, M. Wattenberg, and M. M. McKeon, "The hidden order of wikipedia," in *Proceedings of the 2nd international conference on Online communities and social computing*, ser. OCSC'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 445–454.

[14] K. Panciera, A. Halfaker, and L. Terveen, "Wikipedians are born, not made: a study of power editors on wikipedia," in *Proceedings of the ACM 2009 international conference on Supporting group work*, ser. GROUP '09. New York, NY, USA: ACM, 2009, pp. 51–60.

[15] C. Hess and E. Ostrom, "Introduction: An Overview of the Knowledge Commons," in *Understanding Knowledge as a Commons. From Theory to Practice*, C. Hess and E. Ostrom, Eds. MIT Press, december 2006, pp. 3–26.

[16] O. Arazy, O. Nov, R. Patterson, and L. Yeo, "Information quality in wikipedia: The effects of group composition and task conflict," *Journal of Management Information Systems*, vol. 27, no. 4, pp. 71–98, 2011.

[17] R. Glott, P. Schmidt, and R. Ghosh, "Wikipedia survey – overview of results," UNU-MERIT, Tech. Rep., 2010.

[18] S. Dejean and N. Jullien, "Enrolled since the beginning. an assessement of the wikipedia contributors' behavior regarding their first contribution," *Marsouin working paper*, 2012. [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1980806

[19] H. R. Varian, *Intermediate microeconomics: a modern approach*, 2nd ed. New York, NY: W. W. Norton, 2005.

[20] M. J. Farell, "The measurement of productive efficiency," *Journal of the Royal Statistical Society, Series A*, vol. 120, no. 3, pp. 250–290, 1957.

[21] A. Heshmati, "Productivity growth, efficiency and outsourcing in manufacturing and service industries," *Journal of Economic Surveys*, vol. 17, no. 1, pp. 79–112, 2003.

[22] B. Hecht and D. Gergle, "The Tower of Babel meets Web 2.0: user-generated content and its applications in a multilingual context," in *Proceedings of conference on human factors in computing systems (CHI 2010)*. New York: ACM, 2010.

[23] A. Kittur, E. H. Chi, B. A. Pendleton, B. Suh, and T. Mytkowicz, "Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie," in *CHI 2007*, San Jose, CA, USA, 2007.

[24] F. Ortega and J. M. Gonzalez Barahona, "Quantitative analysis of the wikipedia community of users," in *Proceedings of the 2007 international symposium on Wikis*, ser. WikiSym '07. New York, NY, USA: ACM, 2007, pp. 75–86.

[25] F. Ortega, D. Izquierdo-Cortazar, J. M. Gonzalez-Barahona, and G. Robles, "On the analysis of contributions from privileged users in virtual open communities," in *42th Hawaii International Conference on System Sciences (HICSS)*. Waikoloa, Big Island, Hawaii, USA: IEEE Computer Society, Jan. 2009, p. 110.

[26] A. Lih, "Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource," in *5th International Symposium on Online Journalism*, 2004.

[27] J. Voss, "Measuring wikipedia," in *Proceedings of the International Conference of the International Society for Scientometrics and Informetrics (ISSI)*, ser. 10, Stockholm, 2005.

[28] D. M. Wilkinson and B. A. Huberman, "Assesssing the value of cooperation in wikipedia," *First Monday*, vol. 12, no. 4, Apr. 2007.

[29] S. L. Bryant, A. Forte, and A. Bruckman, "Becoming wikipedian: Transformation of participation in a collaborative online encyclopedia," in *proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. New York, NY, USA: ACM, 2005.

[30] B. A. Kitchenham, "The question of scale economies in software - why cannot researchers agree?" *Information and Software Technology*, vol. 44, no. 1, pp. 13 – 24, 2002.

[31] A. Charnes, W. Cooper, and R. E., "Measuring the efficiency of decision making units," *European Journal of Operational Research*, no. 2, pp. 429–444, 1978.

[32] W. W. Cooper, L. M. Seiford, and J. Zhu, Eds., *Handbook on Data Envelopment Analysis*, 2nd ed. Springer, 2011.

[33] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078–1092, 1984.

[34] R. D. Banker, W. W. Cooper, L. M. Seiford, R. M. Thrall, and J. Zhu, "Returns to scale in different dea models," *European Journal of Operational Research*, vol. 154, no. 2, pp. 345–362, April 2004.

[35] S. Sadiq, "The final frontier: A sas approach to data envelopment analysis," in *SAS Global Forum 2011 Operations Research*, 2011, paper 198-2011.

[36] W. W. Cooper, L. M. Seiford, and K. Tone, *Introduction to data envelopment analysis and its uses: with DEA-solver software and references*. Birkhq̈user, 2006.

[37] Z. van Dijk, "Wikipedia and lesser-resourced languages," *Language Problems and Language Planning*, vol. 33, no. 3, pp. 234–250, 2009.

[38] R. Soekatno and A. Giri, "The indonesian language wikipedia," in *Wikimania05*, 2005. [Online]. Available: http://meta.wikimedia.org/wiki/Transwiki:Wikimania05/Paper-RS1

[39] X. Zhang and F. Zhu, "Group size and incentives to contribute: A natural experiment at chinese wikipedia," *The American Economic Review*, vol. 101, no. 4, pp. 1601–1615, 2011.

[40] A. Kittur, B. Suh, B. Pendleton, and E. H. Chi, "He says, she says: Conflict and coordination in wikipedia," in *CHI 2007*, San Jose, CA, USA, 2007.

[41] B. Stvilia, A. Al-Faraj, and Y. Yi, "Issues of cross-contextual information quality evaluation–the case of arabic, english, and korean wikipedias," *Library & Information Science Research*, vol. 31, no. 4, pp. 232–239, 2009.

[42] D. Lindsey, "Evaluating quality control of wikipedia's feature articles," *First Monday*, vol. 15, no. 4, 2010. [Online]. Available: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2721/2482

[43] L. H. Rector, "Comparison of wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles," *Reference Services Review*, vol. 36, no. 1, pp. 7 – 22, 2008.

[44] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser, "Information quality work organization in wikipedia," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 6, pp. 983–1001, 2008.

[45] D. Lewandowski and U. Spree, "Ranking of wikipedia articles in search engines revisited: Fair ranking for reasonable quality?" *Journal of the American Society for Information Science and Technology*, vol. 62, no. 1, pp. 117–132, 2011.

[46] P. K.-F. Fong and R. P. Biuk-Aghai, "What did they do? deriving high-level edit histories in wikis," in *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, ser. WikiSym '10. New York, NY, USA: ACM, 2010, pp. 2:1–2:10.

[47] D. Anthony, S. W. Smith, and T. Williamson, "Reputation and reliability in collective goods: The case of the online encyclopedia wikipedia," *Rationality and Society*, vol. 21, no. 3, pp. 283–306, 2009.

[48] G. Poderi, "Comparing featured article groups and revision patterns correlations in wikipedia," *First Monday*, vol. 14, no. 5, 2009.

[49] A. Kittur and R. E. Kraut, "Harnessing the wisdom of crowds in wikipedia: quality through coordination," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ser. CSCW '08. New York, NY, USA: ACM, 2008, pp. 37–46.

[50] N. Hara, P. Shachaf, and K. F. Hew, "Cross-cultural analysis of the wikipedia community," *Journal of the American Society of Information Science and Technology*, vol. 61, no. 10, pp. 2097–2108, 2010.

[51] E. S. Callahan and S. C. Herring, "Cultural bias in wikipedia content on famous persons," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 10, pp. 1899–1915, 2011.